# Evaluation of origin-destination matrices based on analysis of data on transport passenger flows

Cite as: AIP Conference Proceedings **2333**, 100002 (2021); https://doi.org/10.1063/5.0041801
Published Online: 08 March 2021

Galina Timofeeva, and Olga Ie

# Evaluation of Origin-Destination Matrices Based on Analysis of Data on Transport Passenger Flows

Galina Timofeeva[1, 2, a)] and Olga Ie[1, 2, b)]

[1)] *Ural State University of Railway Transport, Kolmogorov str., 66, Ekaterinburg, 620034, Russia*
[2)] *Ural Federal University named after First President of Russia B. N. Yeltsin, Mira str., 19, Ekaterinburg, 620002, Russia*

[a)] *Corresponding author: Gtimofeeva@usurt.ru*
[b)] *Electronic mail: Olgaie@mail.ru*

**Abstract.** The problem of reconstructing the origin-destination matrix for transport following fixed routes is studied. Estimates are based on statistics on passengers entering and leaving at each stop. Various models for estimating of origin-destination matrices are analyzed, including the entropy model and the generalized gravitational model. The properties of correspondences of duplicating and complementary routes are taken into account for comparing the models.

## INTRODUCTION

The study of the problem of optimizing transport networks is based on an analysis of existing passenger flows and freight traffic, the forecast of changes in these flows when changing the network structure is also taken into account. Many mathematical models are currently proposed for a detailed analysis of quantitative, structural and parametric changes in traffic flows and loading network elements [1, 2, 3]. The problem of predicting the preferences of passengers when choosing a mode of transport was studied in [4, 5].

The main source of data for modeling is the origin-destination matrix (ODM), which characterizes the total volume of movement of passengers, goods or vehicles between any pairs of points of the transport network. Therefore, the construction of the ODM is really important problem for studying the transport network and traffic flows. The calculation of the origin-destination matrix for urban transport is a complex procedure, including dividing the territory of the city into settlement transport areas, creating a database of the population and the number of places of employment in transport areas, etc.

The problem of reconstructing the origin-destination matrix for transport following fixed routes is divided into several subtasks:

- restoration of the origin-destination matrix for a separate route;
- restoration of the origin-destination matrix for a particular type of transport;
- aggregation of origin-destination matrices taking into account various modes of transport.

One of the conditions for the formation of ODM is the use of objective data on the passengers flows obtained by observations. As a rule, data on the passengers movement in urban vehicles contain observation errors reduced the quality of calculation of the origin-destination matrix. Thus, the calculation algorithm should contain a data analysis and adjustment block.

## METHODS FOR ODM EVALUATING

Statistical methods for evaluating correspondence from observations of flows in the network are most often based on one of the approaches: the maximum likelihood method, the method of moments, the generalized least-squares method and the Bayesian approach.

Methods based on the Markov properties of transport correspondence are less often used to evaluate the correspondence of route transport [6] and to determine missing observations [7]. A method for solving the problem of predicting, estimating and reconstructing the origin-destination matrix based on the Bayesian approach was proposed in [3]. A detailed comparative analysis of three methods for evaluating the origin-destination matrix: the linear programming method, the Bayesian approach, and the time-varying network tomography method was performed in [8].

*Applications of Mathematics in Engineering and Economics (AMEE'20)*
AIP Conf. Proc. 2333, 100002-1–100002-6; https://doi.org/10.1063/5.0041801
Published by AIP Publishing. 978-0-7354-4077-7/$30.00

Evaluation of the matrix of transport correspondence for route transport according to the data on the number of passengers entering and leaving is a special case of the general problem of estimating ODM. Such data determines the matrix of transport correspondence ambiguously. There are various approaches to assessing origin-destination matrices: the gravity model and its generalizations, the entropy model, Markov and semi-Markov models, etc.

Let consider the methods for assessing ODM according to the results of observations of passenger traffic.

## Gravitational Model

One of the first mathematical models used to evaluate correspondence was the gravity model [2, 9]. Such models are constructed by analogy with the law of universal gravitation, according to which all bodies are attracted to each other with a force directly proportional to the product of the masses of these bodies and inversely proportional to the square of the distance between them.

The population or the total volume of the departing (entering) stream is taken as body weight, the distance can be replaced by any cost function associated with movement.

The modified gravitational model is written as follows [10]:

$$\rho_{ij} = k\frac{s_i d_j}{f(c_{ij})}, \quad i \in S, \quad j \in D, \tag{1}$$

where $k$ is the calibration coefficient, $s_i$ is the total volume of people leaving $i \in S$, $S$ is the set of points at which passengers enter, $d_j$ is the total volume of people moving to $j \in D$, $D$ is the set of points at which passengers exit, $f(c_{ij})$ is a gravitational function depending on unit costs of $c_{ij}$ for moving from $i$ to $j$.

Since the distances between the stopping points are relatively small, a non-monotonic dependence of the passenger flow volume on the distance between the stopping points may be used.

The following restrictions must hold, they ensure the conditions for a balanced origin-destination matrix:

$$\sum_{j=1}^{n} \rho_{ij} = s_i, \quad \sum_{i=1}^{m} \rho_{ij} = d_j, \quad \rho_{ij} \geq 0, \quad i \in S, \quad j \in D. \tag{2}$$

Relations (2) can be rewritten as

$$\rho_{ij} = \alpha_i \beta_j s_i d_j f(c_{ij}), \quad i \in S, \quad j \in D. \tag{3}$$

The calibration coefficients $\alpha_i$ and $\beta_j$ may be calculated from the nonlinear system of equations

$$\begin{cases} \alpha_i \sum\limits_{j\in D} \beta_j d_j f(c_{ij}) = 1 \text{ for all } i \in S, \\ \beta_j \sum\limits_{i\in S} \alpha_i s_i f(c_{ij}) = 1 \text{ for all } j \in D. \end{cases} \tag{4}$$

System (4) is consistent only if the total volumes of exit and entry coincide

$$\sum_{i\in S} s_i = \sum_{j\in D} d_j.$$

The choice of the gravitational function $f$ is carried out during the calibration of the model based on a comparison of the calculated data for the model and empirical observations [10].

## PRIMARY DATA PROCESSING

The object of experimental research is Yekaterinburg, whose population is more than 1.5 million people. Public transport is represented by bus, tram and trolley routes.

Research Stages:

1. Primary data processing.

2. Calculation of the origin-destination matrix.
3. Comparison and analysis of correspondence of duplicate routes.

Processing such a number of elements requires very significant computing power. Wolfram Mathematica was used for data analysis and modeling.

The field experiment was chosen as the main method for collecting the initial information. Tellers recorded the number of incoming and outgoing passengers at stops on each route of all types of urban transport. The movement time between stops was also recorded. The source data array contains about 85 000 records, each record includes 12 elements.

There are always registration errors on the stage of the collection of primary information. Therefore, the first stage of data processing is standardization and adjustment of the records. The source data array contains about 2100 names of points of departure (arrival), more than 1000 names are duplicate. This significantly complicates the process of obtaining accurate and objective information about the movement of passengers in the city.

First of all incorrect or incomplete records are detected and excluded. Then, an intellectual analysis of the names of the environmental points was carried out in order to unify the names and eliminate duplicate ones. Thus, we got about 700 points of departure (arrival) instead of 2100.

The following values are known on a route line consisting of $n$ stops:

- $s_i, i = 1, \ldots, n$ is the number of passengers entering at the $i$-th stop, when the bus (trolley bus) moves from the first point to the $n$-th (during a day);
- $d_j, j = 1, \ldots, n$ is the number of passengers leaving at the $j$-th stop on this route (during a day).

The route is linear, not circular, so $d_1 = s_n = 0$.

## CALCULATION OF ODM

The next stage is the calculation of the origin-destination matrix based on this data using the generalized gravitational model (3).

The origin-destination matrix $X$ reflects correspondence between stops, $x_{ij}$ is the number of passengers traveling from the $i$-th point to the $j$-th.

$$X = \{x_{ij}\}, \ \ x_{ij} \geq 0.$$

Natural linear constraints follow from the statement of problems:

$$\sum_{j=i+1}^{n} x_{ij} = s_i, \ \ i = 1, \ldots, n-1, \ \ \sum_{i=1}^{j} x_{ij} = d_j, \ \ j = 2, \ldots, n. \tag{5}$$

Similar relationships hold for the return route.

We assume that the function $f(c_{ij})$ depends on the average time $t_{ij}$ of movement between $i$-th and $j$-th stops. The average transit time between stops is a relatively stable indicator of the transport system in a city and can be estimated based on statistical data.

One of the approximations of the gravitation function has the form

$$f(c_{ij}) = \exp\left(-\gamma t_{ij}^{\theta}\right).$$

In calculating the correspondence of labor migrations researches usually take [11] $\gamma \approx 0.065, \ \theta \approx 1$.

The calibration coefficients $\alpha_i$ and $\beta_j$ are calculated using the following algorithm [10].

**Step 1.** Take matrix $X^{(0)} = \{x_{ij}^{(0)}\}$ as an initial approximation, where

$$x_{ij}^{(0)} = s_i d_j f(t_{ij}) \left( \sum_{l \in D} d_l f(t_{il}) \right)^{-1},$$

here $D$ is the set of points at which passengers exit.

**Step k.** At each step of the method we perform the sequence of operations:

$$\hat{x}_{ij}^{(k)} = \begin{cases} x_{ij}^{(k)} d_j \left( \sum_{i \in S} x_{ij}^{(k)} \right)^{-1}, & \text{if } \sum_{i \in S} x_{ij}^{(k)} > d_j, \\ x_{ij}^{(k)} & \text{otherwise.} \end{cases} \tag{6}$$

$$q_i^{(k)} = s_i - \sum_{j \in D} \hat{x}_{ij}^{(k)}, \quad r_j^{(k)} = d_j - \sum_{i \in S} \hat{x}_{ij}^{(k)}; \tag{7}$$

$$x_{ij}^{(k+1)} = \hat{x}_{ij}^{(k)} + q_i^{(k)} r_j^{(k)} f(t_{ij}) \left( \sum_{l \in D} r_l^{(k)} f(t_{il}) \right)^{-1}. \tag{8}$$

The iterative process of calculating the origin-destination matrix continues until the following equations are satisfied with a sufficient degree of accuracy

$$\sum_{i \in S} x_{ij}^{(k)} = d_j \quad \text{for all } j \in D. \tag{9}$$

The following residual function

$$W(x^{(k)}) = \sum_{j \in D} \left( \sum_{i \in S} x_{ij}^{(k)} - d_j \right)^2.$$

is considered to indicates the fulfilment of equalities (9) with a sufficient degree of accuracy.

The calculations showed a high convergence rate of the algorithm, as a rule an acceptable solution was found in no more than 8 iterations. Figure 1 shows a typical graph of the change in the residual function $W(x^{(k)})$ depending on the iteration number $k$.

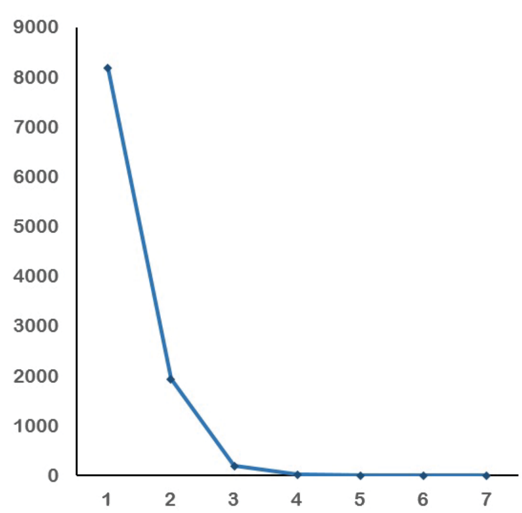


**FIGURE 1.** Dependence of the residual function on the number of iterations.

As a result of applying the described algorithm, we obtain the origin-destination matrix, which is a solution of the gravitational model (1)–(2) for a given set of observations $\{s_i, i \in S\}$, $\{d_j, j \in D\}$.

## COMPARISON OF CORRESPONDENCE FOR DUPLICATE ROUTES

At the next stage of the study, origin-destination matrices for different modes of transport on duplicate routes are compared. The synthetic approach proposed in [1] was used to calculate and analyze ODM for duplicate routes.

A general origin-destination matrix is calculated on the base of the gravitation model (1) using the described algorithm.

Firstly, we find the total passenger traffic data $s_i$, $d_j$

$$s_i = \sum_m \sum_{j \in D} x_{ij}^{(m)}, \quad i \in S, \quad d_j = \sum_m \sum_{i \in S} x_{ij}^{(m)}, \quad j \in D, \tag{10}$$

and general trip costs

$$f(c_{ij}) = \sum_m \exp\left(-\gamma t_{ij}^{(m)}\right), \tag{11}$$

where $m$ is the number of the transport type.

Then values $x_{ij}$ are computed using the algorithm (6)–(8).

The calculated values $x_{ij}$ are substituted into equality (3) instead of $\rho_{ij}$ and products of coefficients $\mu_{ij} = \alpha_i \beta_j$ for all $i \in S$, $j \in D$ for general origin-destination matrix are worked out. The products are determined unambiguously, that is, they are the same for different types of transport and for total passenger flow.

After that, the splitting of correspondences by type of transport is calculated

$$x_{ij}^{(m)} = \mu_{ij} s_i d_j \exp\left(-\gamma t_{ij}^{(m)}\right),$$

where $s_i$, $d_j$ are satisfied to equalities (10).

The fare is the same for all types of urban transport, and the gravity function $f(c_{ij})$ depends only on the average travel time. Therefore, we assume that the travel time is the main factor affecting the values of correspondences. A statistical analysis of the travel time of trolleybuses and buses along the duplicate part of routes was carried out. The bus travel time is on average less than a trolley bus (see Fig. 2).

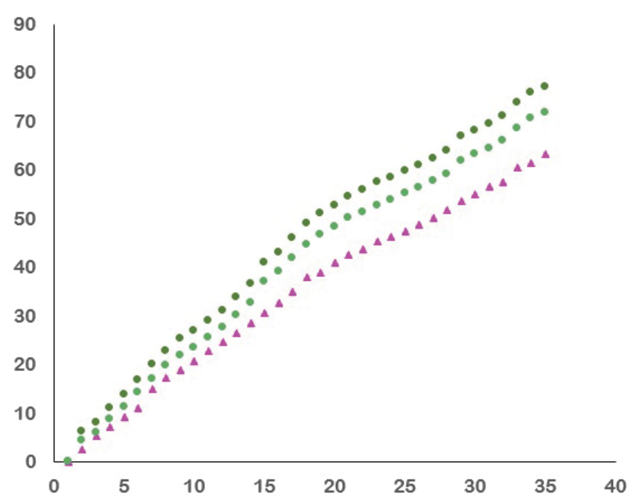


**FIGURE 2.** Pink markers indicate the average travel time for bus; green markers show lower and upper boundary of the confidence interval for the trolleybus travel time; the horizontal axis represents the stop number.

Despite the trolleybus moves on average longer the corresponding passenger flow values are lower for the bus. Therefore, we assume that travel time is not the only determining factor that affects the choice of a passenger.

## CONCLUSION

Origin-destination matrices were studied on the basis of the gravitational model for various types of urban transport in Yekaterinburg. An iterative algorithm for calculating the origin-destination matrix showed a high rate of convergence to the solution, and balanced matrices were obtained in less than 8 iterations. A comparative analysis of the correspondence of various modes of transport on duplicating routes was carried out using the synthetic approach.

The obtained results can be used to adjust the routes of urban passenger transport. The difference in travel time by different modes of transport does not result in significant difference in origin-destination matrices. Thus, the cost function depending only on time costs does not give accurate realistic results. We need to include in the model other factors affecting correspondence or use other types of models.

## ACKNOWLEDGMENTS

The investigation is supported by the Russian Federal Program in the framework of the project ”Optimization of the transport network structure and transport services taking into account the analysis of the networks structure and forecasting demand based on the stochastic model of consumer preference” No AAAA 20-120042190035-7.

## REFERENCES

1. L. G. W. Juan de Dios Ortuzar, *Modelling transport. 4th Edition* (Wiley, 2011).
2. A. M. Voorhees, “A general theory of traffic movement,” Springer **40**, 1105–1116 (2013).
3. A. Ehlert, M. Bell, and S. Grosso, “The optimisation of traffic count locations in road networks,” Transportation Research. Part B. , 460–479 (2006).
4. G. Timofeeva, A. Martynenko, and D. Zavalishchin, “Probabilistic modeling of passengers and carriers preferences via bicriterial approach,” 17th IFAC Workshop on Control Applications of Optimization (CAO 2018), IFAC-PapersOnLine **51 (32)**, 496–498 (2018).
5. G. Timofeeva, “Investigation of the mathematical model of passenger preferences,” AIP Conference Proceedings Applications of Mathematics in Engineering and Economics (AMEE’19). 2172 , 080008 , 6 (2019), DOI:10.1063/1.5133566.
6. B. Li, “Markov models for bayesian analysis about transit route origin-destination matrices,” Transportation Research Part B: Methodological **43(3)**, 301–310 (2009).
7. T. Morimura, T. Osogami, and T. Ide, “Solving inverse problem of markov chain with partial observations,” Advances in Neural Information Processing Systems , 1655—1663 (2013).
8. A. Medina, N. Taft, K. Salamatian, and at al., “Traffic matrix estimation: existing techniques and new directions,” Computer Communication Review. Proc. of the SIGCOMM Conf. , 161–174 (2002).
9. G. A. Carrothers, “An historical review of the gravity and potential concepts of human interaction,” J. American Instit. Planners **22**, 94–102 (1956).
10. G. A. Arrowsmith, “A behavioural approach to obtaining a doubly constrained trip distribution model,” Operational Research Quarterly **24**, 101–111 (1973).
11. E. S. Sheppard, “Gravity parameter estimation,” Geographical Anal **11**, 120–132 (1979).
12. B. Li, “Bayesian inference for origin-destination matrices of transport networks using the em algorithm,” Technometrics **4(47)**, 399–408 (2005).